

# Likelihood transform: making optimization and parameter estimation easier

Yan Wang

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Callinstraße 38, 30167 Hannover, Germany\*

Parameterized optimization and parameter estimation is of great importance in almost every branch of modern science, technology and engineering. A practical issue in the problem is that when the parameter space is large and the available data is noisy, the geometry of the likelihood surface in the parameter space will be complicated. This makes searching and optimization algorithms computationally expensive, sometimes even beyond reach. In this paper, we define a likelihood transform which can make the structure of the likelihood surface much simpler, hence reducing the intrinsic complexity and easing optimization significantly. We demonstrate the properties of likelihood transform by apply it to a simplified gravitational wave chirp signal search. For the signal with an signal-to-noise ratio 20, likelihood transform has made a deterministic template-based search possible for the first time, which turns out to be 1000 times more efficient than an exhaustive grid-based search. The method in principle can be applied to other problems in other fields as the spirit of parameterized optimization and parameter estimation problem is the same.

*Introduction*—Parameterized optimization and parameter estimation is a general important problem in almost every branch of modern science, technology and engineering [1–5]. The general problem can be described as follows. Let us denote  $\theta = (\theta^1, \theta^2, \dots, \theta^k) \in \mathcal{P}$  as the parameters to be estimated, where  $\mathcal{P}$  is the  $k$ -dimensional parameter space. The figure of merit  $\mathcal{F}(\theta, \mathbf{x}) \equiv \mathcal{M} \cdot (\theta, \mathbf{x})$  is usually some functional of the parameters  $\theta$  and the data  $\mathbf{x}$  (e.g. measurement data from experiments). In practice, the functional  $\mathcal{M}$  can be viewed as a set of operations on the parameters  $\theta$  and the data  $\mathbf{x}$ . The goal is to find the best estimate  $\theta_*$  that maximizes the functional  $\mathcal{F}(\theta_*, \mathbf{x})$ . Since minimization can be cast into maximization by just adding a minus sign, we will focus only on maximization problems. Also, for brevity, we omit  $\mathbf{x}$  and denote  $\mathcal{F}(\cdot, \mathbf{x})$  as  $\mathcal{F}(\cdot)$  from now on. The mapping  $\mathcal{M} \cdot (\cdot, \mathbf{x})$  from  $\theta$  to  $\mathcal{F}$  defines a hypersurface in the  $k$ -dimensional parameter space  $\mathcal{P}$ . Hereafter, we refer it as the *likelihood surface* in general, although sometimes it does not have to be related to the likelihood.

By definition, the likelihood surface should peak at the best estimate  $\theta_*$ . When the peak is broad and smooth, there are less structures in this region. Hence, the search is relatively easy and the resolution is poor (i.e. the error bar in the estimate is large). When the peak is sharp and narrow, the resolution is high, but the search is much more difficult. In general, the structure of the likelihood surface determines the difficulty of the optimization problem. By modifying the surface structure, we may alter the innate difficulty of the problem.

*Likelihood transform*—We introduce a set of functionals  $\mathcal{Y}_\sigma$  acting on the mapping functional  $\mathcal{M}$ , where  $\sigma \in \mathcal{Q}$  can be either a scalar variable or a set of variables.

$$\begin{aligned} \mathcal{F}_\sigma(\cdot) &\equiv \mathcal{Y}_\sigma \cdot [\mathcal{M} \cdot (\cdot, \mathbf{x})] \\ &= (\mathcal{Y}_\sigma \cdot \mathcal{M}) \cdot (\cdot, \mathbf{x}), \end{aligned} \quad (1)$$

By varying  $\sigma$ , we obtain a set of modified likelihood surfaces  $\mathcal{F}_\sigma(\cdot)$ . We want to find a proper set of functionals  $\mathcal{Y}_{\sigma(l)}$ , where the index  $l \in [0, l_*] \subset \mathbb{R}$ , such that as  $l$

running from  $l_*$  to 0,  $\mathcal{Y}_{\sigma(l)}$  modifies the sharp narrow peak at the best estimate  $\theta_*$  gradually (or continuously) to broader and smoother hills. We require that  $\mathcal{Y}_{\sigma(l_*)}$  is a unity functional, i.e.  $\mathcal{F}_{\sigma(l_*)}(\cdot) = \mathcal{F}(\cdot)$ . When such a proper set of functionals is identified, one can search on the broadest and smoothest likelihood surface,  $\mathcal{F}_{\sigma(0)}(\cdot)$ , since its (local) maximum  $\theta_\sigma$  is usually easiest to find. Notice that this maximum  $\theta_\sigma$  needs not to be the global maximum on the modified likelihood surface and it needs not to be exactly at  $\theta_*$ . However, as  $l$  going from 0 to  $l_*$ ,  $\theta_\sigma$  should gradually converge to  $\theta_*$  due to the continuity of the transform. This means after identifying the maximum  $\theta_{\sigma(0)}$  in the smoothest likelihood surface, the transform  $\mathcal{Y}_{\sigma(l)}$  can help lead us to the best estimate  $\theta_*$ .

Following the conventions [6] used by the gravitational wave (GW) data analysis community, it is convenient to define the inner product of two time series  $a(t), b(t)$  as below

$$\langle a|b \rangle = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df, \quad (2)$$

where  $\tilde{a}(f), \tilde{b}(f)$  are the Fourier transforms of  $a(t)$  and  $b(t)$ .  $S_n(f)$  is the so-called two-sided power spectral density of Gaussian noise, usually defined as  $E[\tilde{n}^*(f')\tilde{n}(f)] = S_n(f)\delta(f - f')$ .

We denote the normalized GW waveform with parameters  $\theta$  by  $h(\theta, t)$ , thus  $\langle h(\theta)|h(\theta) \rangle = 1$ . The measured data  $x(t)$  containing a GW signal with parameters  $\theta_*$  and Gaussian noise  $n(t)$  can be expressed as  $x(t) = Ah(\theta_*) + n(t)$ , where  $A$  is the strength of the signal.

The figure of merit is the signal-to-noise ratio (SNR)

$$\begin{aligned} \mathcal{F}(\theta) &= \mathcal{M} \cdot (\theta, \mathbf{x}) \\ &\equiv \langle x|h(\theta) \rangle \\ &= \int_{-\infty}^{\infty} \frac{\tilde{x}^*(f)\tilde{h}(\theta, f)}{S_n(f)} df. \end{aligned} \quad (3)$$

Although  $\mathcal{F}(\theta)$  here is not the likelihood, it is directly related to the likelihood  $\mathcal{L}(\theta) \propto \exp[\mathcal{F}(\theta)^2/2]$ . The func-

tional  $\mathcal{M}$  can be interpreted as two operations: first, to generate a waveform with parameters  $\theta$ ; second, to calculate the inner product of this waveform and the data  $\mathbf{x}$ . Usually,  $\mathcal{F}(\cdot)$  has a shape narrow peak at the best estimate  $\theta_*$ . As an example, we define the functionals  $\mathcal{Y}_\sigma$  as convolution operators

$$\begin{aligned}\mathcal{F}_\sigma(\theta) &= (\mathcal{Y}_\sigma \cdot \mathcal{M}) \cdot (\theta, \mathbf{x}) \\ &\equiv \langle x | \mathcal{K}_\sigma \star h(\theta) \rangle \\ &= \mathcal{K}_\sigma \star \mathcal{F}(\theta),\end{aligned}\quad (4)$$

where  $\mathcal{K}_\sigma$  is the kernel function. The last equality is because convolution is a linear operation and  $\mathcal{F}(\theta)$  is linear in  $h(\theta)$ . Since the convolution can be viewed as a smoothing functional, the modified likelihood surface  $\mathcal{F}_\sigma(\cdot)$  is smoother than the original one. For brevity's sake, we temporarily assume  $\theta$  is a scalar parameter and choose the kernel function as one-dimensional Gaussian function  $\mathcal{K}_\sigma = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{\theta^2}{2\sigma^2})$ . The argument below can be trivially generalized to multi-dimensional case. When  $\sigma$  is large, the kernel is a very broad Gaussian function, hence making the likelihood surface  $\mathcal{F}_\sigma(\cdot)$  very smooth. As  $\sigma$  decays to 0,  $\mathcal{F}_\sigma(\cdot)$  converges to the original likelihood surface  $\mathcal{F}(\cdot)$ . Notice that when  $\sigma \rightarrow 0$ , we have  $\mathcal{K}_\sigma \rightarrow \delta(\theta)$ . In practice, we can set  $\sigma(l) = \sigma(0)(l_* - l)/l_*$ . As  $l$  goes from 0 to  $l_*$ ,  $\mathcal{F}_{\sigma(l)}(\cdot)$  evolves gradually from very smooth modified likelihood surface to the original likelihood surface.

From another point of view,  $\mathcal{K}_\sigma \star h(\theta)$  in Eq. 4 is just a weighted average of many waveforms. Since waveforms with similar parameters are correlated, by using a summation of nearby waveforms one can smooth the original likelihood surface. As the number of averaged waveforms goes to 1, the modified likelihood surface converges to the original likelihood surface.

*Applications*—Likelihood transform  $\mathcal{Y}_{\sigma(l)}$  can gradually modify the likelihood surface, hence changing the intrinsic complexity of the optimization problem. In the meantime, it retains the relation between the modified likelihood surfaces and the original likelihood surface. Therefore, likelihood transform can be used in many ways. For instance, it can accelerate stochastic optimization methods, such as Markov chain Monte Carlo [7, 8], particle swarm optimization [7, 9], genetic algorithm [7, 10, 11], etc. It can help design hierarchical search algorithms. In some circumstances, it can even make a deterministic search possible.

Likelihood transform is different from simulated annealing [12], which is also a technique to accelerate stochastic optimization algorithms. In the following, we will compare the two. Simulated annealing employs a temperature parameter  $\mathcal{T}$  to heat the likelihood surface from  $\mathcal{L}(\theta) \propto \exp[\mathcal{F}(\theta)^2/2]$  to  $\exp[\mathcal{F}(\theta)^2/2\mathcal{T}]$ . As the stochastic search algorithms proceed, the temperature cools down gradually. Therefore, the heuristics can escape from local maxima more easily and explore

the whole parameter space more thoroughly, hence being accelerated. As an example, we simulated a sinusoidal signal with only one parameter  $\omega = 0.2 \text{ rad/s}$  buried in white Gaussian noise. The SNR was 9. Fig. 1 (a) shows how simulated annealing gradually modifies the likelihood surface (or more rigorously the SNR surface). As seen from the figure, the likelihood surface is less spiky at high temperatures. Notice that the number of local maxima (including the global maximum) is the same at all temperatures, and the locations of these maxima are unchanged.

As for likelihood transform, it modifies the likelihood as  $\exp[\mathcal{F}_{\sigma(l)}(\theta)^2/2]$ . For simplicity, we choose the functional  $\mathcal{Y}_\sigma$  to be convolution operators with a Gaussian kernel. We then apply it to the same simulated data. The modified likelihood surface  $\mathcal{F}_{\sigma(l)}(\theta)$  are shown in Fig. 1 (b). Notice that both the number of local maxima and their locations are changed by the likelihood transform. In addition, the likelihood surfaces are smoother with less structures comparing to the cases of simulated annealing. These features of likelihood transform may help accelerate the stochastic optimization algorithms more efficiently.

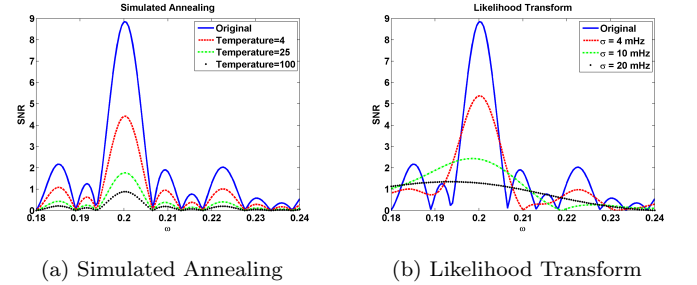


FIG. 1: A comparison of simulated annealing and likelihood transform.

Next, we applied likelihood transform to a toy model in gravitational wave data analysis. Although it is a somewhat simplified model, it captures important features of the general problem and can demonstrate the method in a more general (less background-dependent) way.

Suppose the normalized gravitational wave chirp signal is in the following form

$$h(\theta, t) = \sqrt{\frac{2}{T}} \cos[\omega_0 \alpha_0(t) + \omega_1 \alpha_1(t)], \quad (5)$$

where  $\theta = (\omega_0, \omega_1)$  are the two parameters to be estimated, the two time functions are defined as  $\alpha_0(t) \equiv t - \frac{T}{2}$  and  $\alpha_1(t) \equiv (t - \frac{T}{2})^2$ ,  $T$  is the total observation time, which we choose to be 5120 seconds in the simulation. We inject a signal with parameters  $\omega_0 = 0.0628 \text{ rad/s}$ ,  $\omega_1 = 6.136 \times 10^{-6} \text{ rad/s}^2$ . Notice that  $\omega_0$  is an angular frequency. The searching parameter ranges for  $\omega_0$  and  $\omega_1$  are  $(\omega_0^{\text{low}}, \omega_0^{\text{upp}}) = (1.2 \text{ mrad/s}, 0.126 \text{ rad/s})$  and

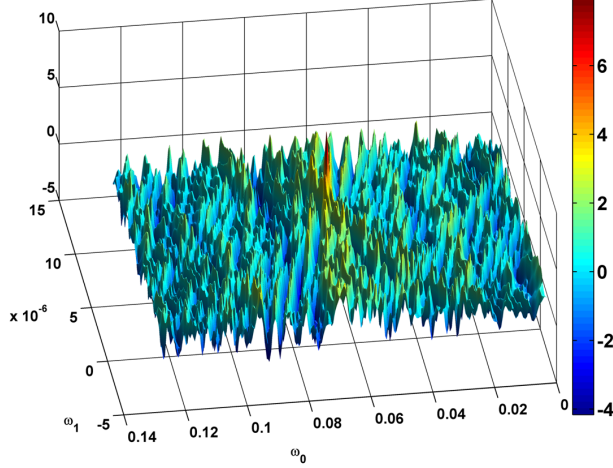


FIG. 2: The original likelihood surface  $\mathcal{F}(\omega_0, \omega_1)$ . It peaks at the true signal parameter with an optimal SNR 8. It is very fluctuant. Optimal template layout requires 69620 templates.

$(\omega_1^{\text{low}}, \omega_1^{\text{upp}}) = (-3.07 \times 10^{-6} \text{ rad/s}^2, 1.23 \times 10^{-5} \text{ rad/s}^2)$  respectively.

As an example, we use convolution operators as the functionals  $\mathcal{Y}_\sigma(l)$  and assume the kernel function to be a Gaussian function with diagonal covariance. Then, we have

$$\begin{aligned} H_{\sigma(l)}(\theta) &= \mathcal{K}_{\sigma(l)} \star h(\theta), \\ &= \int \int \mathcal{K}[\omega_0 - \omega'_0, \omega_1 - \omega'_1, \sigma_0(l), \sigma_1(l)] \\ &\quad \cdot h(\omega'_0, \omega'_1) d\omega'_0 d\omega'_1, \\ &= h(\theta) e^{-\frac{1}{2}[\sigma_0(l)^2 \alpha_0(t)^2 + \sigma_1(l)^2 \alpha_1(t)^2]}, \end{aligned} \quad (6)$$

$$\mathcal{F}_{\sigma(l)}(\theta) = \langle x | H_{\sigma(l)}(\theta) \rangle. \quad (7)$$

We set  $l_* = 1$  and choose  $(\sigma_0(l), \sigma_1(l)) = (1-l)(\omega_0^{\text{upp}} - \omega_0^{\text{low}}, \omega_1^{\text{upp}} - \omega_1^{\text{low}})$  to be a fraction of the entire searching parameter range. We will see how this parameter  $l$  can modify the likelihood surface and adjust the difficulty of the optimization problem. In general, the difficulty of the search can be very well described by the required number of templates for a certain mismatch by template-based search. Following conventions, we set the mismatch to be 0.03. By calculating the metric of the likelihood surface on the parameter space [13], it's straightforward to estimate the number of templates required by optimal layout (we choose rectangular layout here). When  $l = l_* = 1$ , we have the original likelihood surface  $\mathcal{F}(\omega_0, \omega_1)$  shown in Fig. 2. On this likelihood surface, the optimal layout requires 69,620 templates. This large required number is due to the noise-like features of the likelihood surface.

The structure of the likelihood surface can be greatly simplified through the likelihood transform. Fig. 3 (a)-(d) shows the several transformed likelihood surfaces

$\mathcal{F}_{\sigma(l)}(\omega_0, \omega_1)$  with different values of  $l$ . When  $l = 3/4$ , the modified likelihood surface is very smooth. It is extremely simple to characterize the structure of this likelihood surface or find its maximum. As  $l$  increases, more and more structures appear on the likelihood surface  $\mathcal{F}_{\sigma(l)}(\omega_0, \omega_1)$ . It gradually converges to the original likelihood surface  $\mathcal{F}(\omega_0, \omega_1)$ . These figures show how the difficulty of the optimization problem can be modified by the likelihood transform. More precisely, we have calculated the required number of templates for different values of  $l$  in Fig. 4.

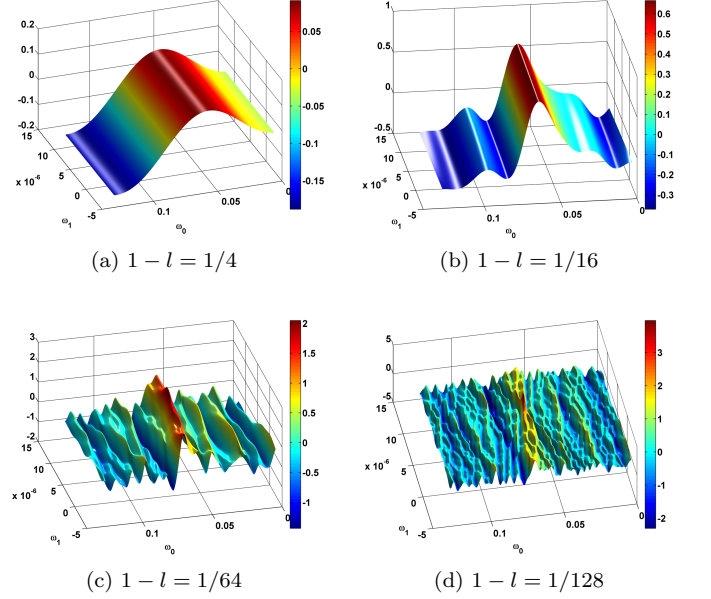


FIG. 3: The modified likelihood surfaces  $\mathcal{F}_{\sigma(l)}(\omega_0, \omega_1)$  after likelihood transforms.

For  $1-l > 0.1$ , the 0.03-mismatch rule gives an error rectangular, which is comparable to the area of the entire search parameter space. However, the error rectangular may have very different shape from the search parameter space. Therefore, in  $1-l > 0.1$  region, the required number of templates shown in Fig. 4 only serves as a rough estimate of the complexity of the modified likelihood surface. In the more interesting  $1-l < 0.1$  region, the dependence of the required number on  $1-l$  roughly follows a power law. The required number decreases rapidly in this region, hence the difficulty of search decreases rapidly.

These features of likelihood transform can potentially help the optimization algorithms. For example, it may help in the design of efficient hierarchical algorithms to search for GW signals.

In the following, we will show that in some cases the likelihood transform can even make deterministic optimization methods such as Newton's method possible. Hence, the search algorithm will be much more efficient.

In the neighbourhood of a (local) maximum  $\theta_{\sigma(l)}$  on

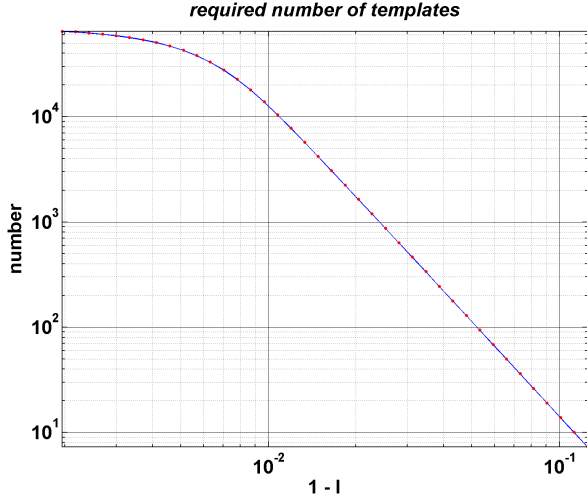


FIG. 4: Number of templates required by the optimal layout for different values of  $l$ . Notice that the original case  $l = 1$  is not plotted here.

modified likelihood surface  $\mathcal{F}_{\sigma(l)}(\cdot)$ , the geometry can be described by a Taylor series

$$\begin{aligned} \mathcal{F}_{\sigma(l)}(\theta_{\sigma(l)} + \Delta\theta) &= \mathcal{F}_{\sigma(l)}(\theta_{\sigma(l)}) \\ &+ \frac{1}{2} \frac{\partial^2 \mathcal{F}_{\sigma(l)}}{\partial \theta^\mu \partial \theta^\nu} \Delta\theta^\mu \Delta\theta^\nu + \mathcal{O}(\Delta\theta^3). \end{aligned} \quad (8)$$

where we have assumed the Einstein summation convention. The first derivative vanishes and the modified Fisher information matrix  $I_{\mu\nu}^{\sigma(l)} = -\frac{\partial^2 \mathcal{F}_{\sigma(l)}}{\partial \theta^\mu \partial \theta^\nu} |_{\theta=\theta_{\sigma(l)}}$  is positive definite due to the fact that  $\mathcal{F}(\theta_{\sigma(l)})$  is a maximum stationary point. Notice that, when  $l = l_*$ , the Taylor expansion is on the original likelihood surface around the best estimate  $\theta_*$ , and  $I_{\mu\nu} = I_{\mu\nu}^{\sigma(l_*)} = -\frac{\partial^2 \mathcal{F}}{\partial \theta^\mu \partial \theta^\nu} |_{\theta=\theta_*}$  is the Fisher information matrix at the best estimate  $\theta_*$ . For each modified likelihood surface  $\mathcal{F}_{\sigma(l)}(\cdot)$ , there exists a neighbourhood  $\mathcal{B}_l \subseteq \mathcal{P}$  of  $\theta_{\sigma(l)}$  where the geometry of the likelihood surface can be approximated by a quadratic form quite well (say, the percentage error caused by higher order term is less 1%). According to our design, the smaller the  $l$  the smoother the modified likelihood surface  $\mathcal{F}_{\sigma(l)}(\cdot)$ , hence the larger the neighbourhood  $\mathcal{B}_l$ . Sometimes,  $\mathcal{B}_0$  can be as large as the entire parameter space  $\mathcal{P}$ .

Starting from any point  $\theta'_l \in \mathcal{B}_l$  on modified likelihood surface  $\mathcal{F}_{\sigma(l)}(\cdot)$ , one can easily find the best estimate  $\theta_{\sigma(l)}$  via some deterministic local-search algorithms. For instance, by neglecting higher order terms in Eq. 8 and differentiating both sides with respect to  $\theta^\nu$ , we get

$$\frac{\partial \mathcal{F}_{\sigma(l)}(\theta'_l)}{\partial \theta^\nu} = \frac{\partial^2 \mathcal{F}_{\sigma(l)}}{\partial \theta^\mu \partial \theta^\nu} \Delta\theta^\mu. \quad (9)$$

Thus, we calculate the best estimate in just one step

$$\begin{aligned} \theta_{\sigma(l)}^\mu &= \theta'_l{}^\mu - \Delta\theta^\mu \\ &= \theta'_l{}^\mu - \left[ \frac{\partial^2 \mathcal{F}_{\sigma(l)}}{\partial \theta^\mu \partial \theta^\nu} \right]^{-1} \frac{\partial \mathcal{F}_{\sigma(l)}(\theta'_l)}{\partial \theta^\nu}, \end{aligned} \quad (10)$$

where  $\frac{\partial^2 \mathcal{F}_{\sigma(l)}}{\partial \theta^\mu \partial \theta^\nu}$  is constant in  $\mathcal{B}_l$ , so it can be calculated at  $\theta'_l$ . Observe that as  $l$  gradually runs from 0 to  $l_*$ ,  $\mathcal{B}_l$  shrinks smoothly. Also, since  $\mathcal{B}_l$  is roughly a quadratic region,  $\theta_{\sigma(l)}$  should be near the center of  $\mathcal{B}_l$ . So, there must exist a smaller region  $\mathcal{B}_{l_1}$  (with  $l_1 > l$ ) which contains  $\theta_{\sigma(l)}$  in it. One can take  $\theta_{\sigma(l)}$  as the starting point in  $\mathcal{B}_{l_1}$  and repeat Eq. 10 to calculate the best estimate  $\theta_{\sigma(l_1)}$  on  $\mathcal{F}_{\sigma(l_1)}(\cdot)$ . By iterating the above process, one will find the best estimate  $\theta_*$  on the original likelihood surface  $\mathcal{F}(\cdot)$ .

Usually, we need to study the properties of the neighbourhood  $\mathcal{B}_l$  in order to design an efficient deterministic algorithm. However, in some cases, likelihood transform can change the likelihood surface to be so smooth and regular that we can simply choose a set of  $\mathcal{V}_\sigma$  to perform a deterministic search. As an example, we still use the waveform model introduced in the last subsection and set the SNR to 20. Six transformed likelihood surfaces are shown in Fig. 5. Notice that the global maxima on these surfaces are normalized to 1. We start from 10 points in the parameter space uniformly sampled in  $\omega_0$  with random  $\omega_1$ . Then, we calculate the values of  $\mathcal{F}_\sigma(\theta)$  at these 10 points on the smoothest transformed likelihood surface. The maximum among these 10 points is set as the initial location for the Newton's method with 10 iterations. Fig. 6 shows the simulation result of this deterministic algorithm. After 7 iterations, this algorithm converges to the location of the global maximum of the original likelihood surface. In this process, we have only used a few tens of templates. Comparing to 69,620 templates required by a grid-based search algorithm, the deterministic algorithm is about 1,000 times more efficient.

*Conclusion*—We have introduced the likelihood transform as a general tool to make optimization and parameter estimation easier. The likelihood transform can gradually transform the likelihood surface to a smoother shape with less complex structure. On these modified likelihood surfaces, the local and global maxima are much easier to find. Since these modified likelihood surfaces are directly related to the original likelihood surface by the likelihood transform, one can find the global maximum of the original likelihood surface more efficiently based on knowledge of the transformed likelihood surfaces. We have shown the possibility to use likelihood transform to accelerate stochastic optimization methods. Compared to simulated annealing, likelihood transform gives indications that it would accelerate the heuristics more efficiently. We applied likelihood transform to a GW data analysis problem with a toy waveform model. Simulation results show that likelihood transform can manipulate the structure of the original likelihood sur-

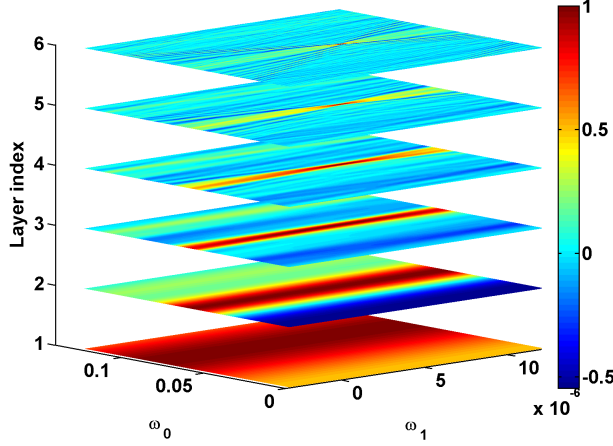


FIG. 5: A plot of six transformed likelihood surfaces  $\mathcal{F}_{\sigma(l)}$ .

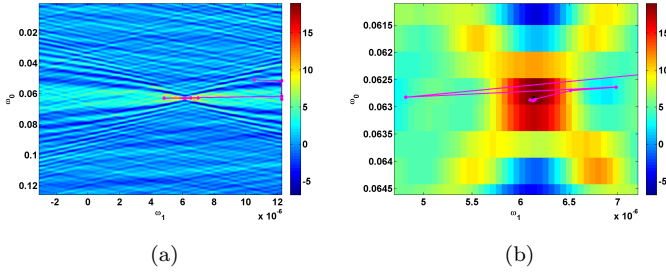


FIG. 6: A deterministic search with the help of likelihood transform. The pink points identify the trajectory of  $\theta_{\sigma(l)}^{\mu}$ . (b). is a zoom-in version of (a).

face, hence allowing it to combine with and accelerate a hierarchical search. We have also shown that for the toy waveform model with SNR= 20, likelihood transform make a deterministic search possible, which turns out to be 1,000 times more efficient than the exhaustive grid-based search for GW signals. With the help of likelihood

transform, a template-based deterministic search for GW signals is shown to be possible for the first time.

In this article, we have only considered linear functionals, or more specifically, convolutions with Gaussian kernels with uncorrelated covariances. In the future, we will study other linear functionals and even nonlinear functionals  $\mathcal{Y}_{\sigma}$ , which would potentially exhibit better properties.

The author would like to thank S. Babak, B. Krishnan and R. Prix for useful comments. The author was partially supported by DFG Grant No. SFB/TR 7 Gravitational Wave Astronomy. The author would also like to thank German Research Foundation for funding the Cluster of Excellence QUEST-Center for Quantum Engineering and Space-Time Research.

\* yan.wang@aei.mpg.de

- [1] E. Poisson and C. M. Will, Phys. Rev. D **52**, 848-855 (1995).
- [2] K. Tamura et al., Molecular Biology and Evolution, Volume 28, Issue 10, Pp. 2731-2739 (2011).
- [3] T. Bollerslev and J.M. Wooldridge, Econometric Reviews, Volume 11, Issue 2 (1992).
- [4] R.O. Schmidt, Antennas and Propagation, IEEE Transactions on Volume 34, Issue 3 (1986).
- [5] S. Johansen and K. Juselius, Oxford Bulletin of Economics and Statistics, Volume 52, Issue 2, pages 169C210 (May 1990).
- [6] P. Jaranowski and A. Królak, Living Rev. Relativity, 15, (2012), 4, <http://www.livingreviews.org/lrr-2012-4>.
- [7] Y. Wang, Y. Shang, and S. Babak, Phys. Rev. D **86**, 104050 (2012).
- [8] T. B. Littenberg, and N. J. Cornish, Phys. Rev. D **80**, 063007 (2009).
- [9] Y. Wang, and S. D. Mohanty, Phys. Rev. D **81**, 063002 (2010).
- [10] J. Crowder, N. J. Cornish, and J. L. Reddinger, Phys. Rev. D **73**, 063011 (2006).
- [11] A. Petiteau, Y. Shang, S. Babak, and F. Feroz, Phys. Rev. D **81**, 104016 (2010).
- [12] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, Science 13 (May 1983): Vol. 220 no. 4598 pp. 671-680.
- [13] B. J. Owen, Phys. Rev. D **53**, 6749-6761 (1996).